

УДК 665.9

АНАЛІЗ МЕТОДІВ ПОШУКУ ФАЙЛІВ-АНАЛОГІВ НА МНОЖЕНІ ЗБЕРЕЖЕНИХ ЗОБРАЖЕНЬ

В.А. МИНКО^{1*}, Л.Б. КАЩЕЄВ²

^{1.} *магістрант кафедри Системного аналізу та інформаційно-аналітичних технологій,
НТУ «ХПІ», Харків, УКРАЇНА*

^{2.} *професор кафедри Системного аналізу та інформаційно-аналітичних технологій,
канд. техн. наук, НТУ «ХПІ», Харків, УКРАЇНА*

**email: victoriamaynko@gmail.com*

Стрімкий розвиток портативної електроніки, зокрема смартфонів та планшетів, і широка доступність стабільного широкосмугового доступу до інтернету та соціальних мереж дали можливість користувачам створювати свій медіа-контент та обмінюватися ним. Причому обсяги переданої та збереженої інформації з плином часу тільки збільшуються. Все це призводить як до того, що користувачам необхідно якимось чином управляти своїми колекціями фотографій, відео та аудіозаписів, так і до того, що інтернет-сервісам і контент-провайдерам необхідно справлятися з усе зростаючим обсягом інформації, що зберігається і передається, як можна більш ефективно.

Зберігання великого обсягу візуальних матеріалів не є проблемою з технічної точки зору. У той же час особливої актуальності набуває завдання забезпечення ефективного змістовного доступу до релевантної інформації в електронних колекціях зображень. Подібна задача може виникнути, наприклад, при реалізації сервісу пошуку зображень за заданим зразком (GoogleImages, TinEye і т.п.), або для пошуку і фільтрації дублікатів зображень, які займають ресурси сервера на веб-порталі, який надає можливість користувачам вільно завантажувати особисті фото і відео.

Ще приклад: одна із сучасних форм роботи музеїв, бібліотек та архівів – формування електронних колекцій зображень, з метою збереження культурної спадщини та її пропагування за рахунок забезпечення віддаленого доступу. Наприклад, при занесенні даних про різноманітні експонати досить зручно робити їх фото та прикріплювати до опису в електронній бібліотеці музею, або й взагалі до загальної системи музеїв світу, яка б дозволяла продивлятися зображення експонатів кожного з них, порівнювати та аналізувати їх завдяки електронним колекціям [1].

Не менш важливий ефективний пошук зображень для підвищення достовірності діагностики при проведенні комп'ютерних досліджень різних захворювань, де зображення об'єкта має максимально збігатися з відповідними формою і розмірами зображеного об'єкта при скороченні обчислювальних витрат.

Виходячи з актуальності даної проблеми та ступеня розвитку різноманітних досліджень в цьому напрямленні, метою даної роботи є

реалізація автоматизованої системи пошуку файлів-аналогів на множені зображень.

У зв'язку з тим, що більша кількість існуючих алгоритмів, призначених для пошуку дублікатів файлів, мають складність $O(N^2)$ – кожне зображення порівнюється одне з одним, то це призводить до великих витрат часу, тим самим, впливаючи на ефективність роботи програм. Рішенням даної проблеми являється спосіб, який базується на попередньому (перед безпосереднім використанням алгоритмів) порівнянні зображень по заданим числовим характеристикам, які можна отримати без додаткових обчислень. До таких характеристик зображень відносяться назва та розмір зображення (ширина та висота), формат та роздільна здатність, співвідношення сторін зображення, дата створення, формати стиснення та інші. Даний спосіб полегшує обробку та порівняння зображень та зменшує складність обчислень до $O(N^2)$.

Саме тому розроблена система має можливість здійснювати попереднє порівняння зображень за наступними критеріям: графічний формат, висота/ширина зображення, розмір та роздільна здатність зображення, Image- та Pixel-формат і розмір друкованого відбитку. Вона працює із зображеннями у таких графічних форматах як .jpg (.jpeg), .png, .bmp, .gif та .ico.

Також слід зазначити, що розроблена система дає можливість користувачеві видаляти зображення, які на його погляд, являються дублікатами. В свою чергу для більш ефективного та зручного перегляду зображень в колекції, реалізована панель із зображеннями, яка складається із трьох вікон і дає можливість одразу переглянути активне зображення, тобто, яке переглядає користувач, та два наступні

Отже, виходячи з того, що знайти алгоритм або програму, які б змогли оцінювати та приймати всі особливості зображення, бути стійкими до модифікацій та швидко опрацьовувати великі обсяги файлів майже не можливо. Тому оптимальним рішенням даної проблеми є попереднє зіставлення зображень по характеристикам зображенням, які є у зображення з моменту його створення та, які не потребують додаткового обчислення. Такий спосіб дозволить зменшити кількість зображень для їх подальшого порівняння за допомогою різноманітних «математичних» алгоритмів та прискорить роботу програм, призначених для пошуку дублікатів.

Список літератури:

1. Паромова Т. А. Исследование схожести изображений с помощью перцептивных хэш-алгоритмов и расстояний Хэмминга / Т. А. Паромова, И. Я. Зеленева, Н. В. Луценко, С. А. Кленачов // Наукові праці Донецького національного технічного університету. Серія: Інформатика, кібернетика та обчислювальна техніка. – 2016. – Вип. 2. – С. 62-68. – Режим доступу: http://nbuv.gov.ua/UJRN/Npdntu_inf_2016_2_9.